Project no.: FP7-ICT-217077
Project full title: Heterogeneous 3-D Perception across Visual Fragments
Project Acronym: EYESHOTS
Deliverable no: D5.4 update
Title of the deliverable: Report on cooperative behaviour in shared workspace.

Date of Delivery:	28 February 2011	
Organization name of lead contractor for this deliverable:	WWU	
Author(s):	R. Volcic and M. Lappe	
Participant(s):	WWU	
Workpackage contributing to the deliverable:	WP5	
Nature:	Report	
Version:	1.0	
Total number of pages:	39	
Responsible person:	Markus Lappe	
Revised by:	P. Fattori	
Start date of project:	1 March 2008	**Duration:** 36 months

Project Co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

This deliverable reports data on allocation of attention, direction of gaze, and movement of the arm of a human cooperation partner. Data were collected in three separate studies. The first two studies are single actor setting experiments, whereas the third study is based on a human-human interaction experiment. We approached the following questions: (a) Can gaze direction changes be used to predict forthcoming relevant locations? (b) Since eye movements are performed constantly, how do human distinguish relevant from irrelevant eye movements? (c) Does an adaptable and cooperative behavior lead to a reduction of attentional resources needed for a successful execution of a task?

Contents

1 Executive summary **2**

2 Introduction **3**

3 Study I: Predictive eye movements in gaze and action observation **3**

3.1 Introduction . . . 3
3.2 Materials and Methods . . . 5
3.2.1 Participants . . . 5
3.2.2 Apparatus . . . 5
3.2.3 Stimuli and Procedure . . . 6
3.2.4 Data Analysis . . . 8
3.3 Results . . . 8
3.3.1 Number of saccades for target identification . . . 8
3.3.2 Timing of target identification . . . 8
3.4 Discussion . . . 10

4 Study II: Potentially purposeful actions divert overt attention **11**

4.1 Introduction . . . 11
4.2 Materials and Methods . . . 14
4.2.1 Participants . . . 14
4.2.2 Apparatus . . . 14
4.2.3 Stimuli and Procedure . . . 14
4.2.4 Data Analysis . . . 17
4.3 Results . . . 18
4.3.1 Influence of congruency and SOA . . . 18
4.3.2 Congruency effects between conditions . . . 20
4.4 Discussion . . . 22

5 Study III: Gaze behavior in cooperative action **26**

5.1 Introduction . . . 26
5.2 Materials and Methods . . . 27
5.2.1 Participants . . . 27
5.2.2 Apparatus . . . 27
5.2.3 Procedure . . . 28
5.2.4 Data Analysis . . . 30
5.3 Results . . . 31

5.3.1 Stereotypical gaze behavior 31
5.3.2 Timing of object directed saccades: comparison between conditions 32
5.4 Discussion 35

References **37**

1 Executive summary

This deliverable describes the research work performed by WWU regarding Task 5.4 (Predicting behavior and cooperation in shared workspace) of Work Package 5 (Human behavior and neural correlates of multisensory 3D representation). Task 5.4 is concerned with the study of the allocation of attention, direction of gaze, and movement of the arm of a human cooperation partner. The goal of this research line is to understand the mechanisms on which the humans abilities of predicting the behavior of a cooperating partner are based. A description of the human's behavior will be then used for prediction of the human behavior for the robot in WP4. By the mechanisms of shared attention the robot will be able to track a human partner's overt attention and predict and react to the partner's actions (Principal Objective 3).

In particular, WWU activities have been focused on the development of two experimental set-ups and, successively, on data collection and data analysis. One experimental set-up is based on a single eye-tracker device to measure eye movements while stimuli, as static images or movies, are presented on a monitor. The second experimental set-up is based on two eye-trackers that allow simultaneous recording of eye movements of two interacting participants in a common task. Moreover, the action space shared by the two participants is also recorded by two head-mounted video-cameras. In this way, eye movements can be superimposed on the video recordings to monitor and study the interaction of gaze behavior, hand movements and object handling.

In the present report, three human behavioral studies are presented. The first two studies are single actor setting experiments, whereas the third study is a human-human interaction experiment.

2 Introduction

Humans and many other species tend to look at things in their surrounding environment that are of immediate interest to them. Individuals who are able to rapidly detect when they are the object of another's attention and who can accurately analyze where another's gaze is directed have considerable adaptive advantage. Equipping a robot with similar capabilities would certainly provide additional adaptability to its behavior that is especially important in human-robot settings, where the robot should adapt its actions to the actions of an active cooperative partner.

In order to develop the understanding of human gaze behavior we focused our research on questions that are directly connected to the needs of a robotic system. First, we approached the question if humans are not only able to accurately judge the direction of gaze, but also predictively react to it. If so, a robotic system that would use humans' gaze information to anticipate certain actions, would actually act in a human-like way by anticipating humans' actions towards targets in space. Second, since humans perform eye movements constantly (on average three eye movements per second), but just a fraction of them are actually purposeful eye movements, we researched how humans distinguish relevant from irrelevant eye movements. This knowledge would have important implications also for the robotic system in reducing the amount of information to which the robot should respond or adapt its behavior. Third, in a human-human interaction study we explored the gaze behavior of both human partners in a simple cooperative task. Moreover, we wanted to estimate the gain of an adaptable and cooperative behavior and see if an adaptable and cooperative interaction actually leads to a reduction in the attentional resources and cognitive load needed in the execution of a task. A robotic system that would allow a similar reduction in the deployment of humans' attentional resources would definitely gain in its usability.

3 Study I: Predictive eye movements in gaze and action observation

3.1 Introduction

Gaze is regarded in many research fields as an important component of non verbal cues. In the field of psychophysics, gaze has been investigated as a particularly important social cue. For instance, Gibson and Pick (1963) measured the relation between the gaze direction toward a participant and the sensation of being looked at. They showed that participants felt not being looked at when the looker's gaze direction was deviated from the nose more than 2.8°. Cline (1967) and Anstis et al (1969) extended Gibson's experiment by measuring the relation between the looker's gaze directions and the participant's judged directions. In these cases, participants judged not only whether they were looked at, but also where the looker was looking. Their general result is that judgments of gaze direction are quite accurate for lookers with frontally oriented heads. Similar results showing that gaze following is quite precise were more recently obtained also by Symons et al (2004) and Bock et al (2008).

Knowing where another person is looking is an important social and cognitive skill that poses the basis of joint attention. An important aspect of joint attention are also the objects on which the looker is attending to. For instance, Schwaninger et al (2005) showed that gaze targets can be determined quite accurately. They presented faces on a computer screen that were fixating one of four invisible target points. The task was to judge where the looker was gazing by placing a cursor on the perceived fixation point using the mouse. They found a general overestimation toward the outside of the actual gaze direction. In a follow-up study, Lobmaier et al (2006) showed that the presence of objects in the attended space also has an impact on gaze interpretation. Objects captured the perceived gaze line, but cues from the eye were also taken into account. Gaze processing is thus biased toward the assumption that a person is looking at an object rather than at an empty space. A visible object is actually likely to be relevant for the observer and it is thus sensible to assume that an observed person will probably attend to an object that is relevant at that moment in time.

To our knowledge, all of the studies on gaze direction were principally focused on the accuracy of the observers' judgments. However, the crucial aspect in joint attention is the ability to predict the focus of the partner's attention. The implication is that observers should be quicker in identifying the target of interest than the partner is in acting on that same target. In the present study, we focused on the participants' ability in predicting the forthcoming relevant locations. In natural hand pointing behavior toward targets of limited spatial extent, gaze always precedes the hand movement. First, the target is fixated and then the hand motor act starts developing towards the target. Being able to predict the end point of the hand movement by correctly interpreting the gaze direction should thus always result in a temporal advantage in the identification of the relevant object.

Participants watched movies of an actor performing a hand pointing movement towards one of several targets. Both the face and the upper-body of the actor were visible. The participants' task was to direct their own gaze as soon as possible towards the target that the actor was going to point to. The gaze behavior of the participants was measured with an eye-tracker. We studied both the effect that the availability of the actor's gazing behavior had on the identification of the correct target, and the effect that the presence or absence of visible targets had on the the identification of the correct location. Therefore, we manipulated the movies in the following way. The actor's eye were either visible or occluded, thus providing or excluding the actor's gazing behavior. And, the targets toward which the actor was pointing were either visible or not, thus providing or excluding the physical presence of the to be pointed locations. On the other hand, the whole movement of the actor's arm from the resting position to the target was always completely visible.

We reasoned that if the ability to use the gaze direction of a partner actually helps predicting the forthcoming relevant locations in peripersonal space, then the participants should identify the correct target more quickly than when the gaze information of the partner is not available. Most importantly, the identification of the correct target should occur before the actor's finger reaches the target.

3.2 Materials and Methods

3.2.1 Participants

Six male participants aged from 22 to 31 year participated in the experiment after providing written informed consent. All had normal or corrected-to-normal vision and were naïve as to the purpose of the experiment. The experimental protocol was conducted according to the Declaration of Helsinki.

3.2.2 Apparatus

In the experiment, the participant sat 54 cm in front of a 20 inch computer monitor (Iiyama Vision Master Pro 514) with a vertical frequency of 100 Hz at a resolution of 1600×1200. The head was stabilized by a chin rest. For stimuli presentation and data collection we used MATLAB with Psychtoolbox (Brainard, 1997; Cornelissen et al, 2002; Pelli, 1997). Eye movements were monitored by the head-mounted Eyelink II system (SR Research Ltd., Mississauga, Ontario, Canada), which samples gaze positions with a frequency of 500 Hz. Viewing was binocular, but only the dominant eye was recorded. Calibration was performed at the beginning of the recorded session and drift correction was performed before each trial to ensure optimal accuracy.

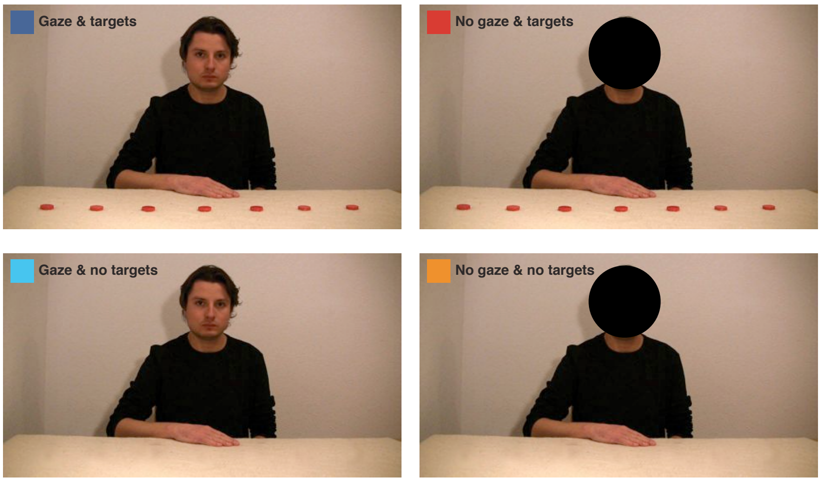

Figure 1: The four experimental conditions measured in the present study that differed on the basis of two factors: (a) the presence of the gazing behavior (top left and bottom left versus top right and bottom right), and (b) the visibility of the targets (top left and top right versus bottom left and bottom right).

3.2.3 Stimuli and Procedure

Stimuli were colored digital movies (960 × 540 pixels) presented at 25 frames/second. Movies were captured with a Canon HV20 high definition camcorder. In each movie, an actor performed a pointing arm movement towards one of six targets positioned on the table in the frontoparallel plane. The actor first gazed at the to be pointed target and the executed the pointing movement with his arm. The duration of each movie was 1600 ms. In all movies, 400 ms after the start of the movie the gaze shift was performed. The arm movement started 240-280 ms after the start of the gaze shift and ended after 600-720 ms depending on the target location. The same movies were presented in four different conditions that differed on the basis of two factors (see Figure 1): (a) the presence of the gazing behavior, and (b) the visibility of the targets. With the first factor the availability of the gazing behavior was manipulated. The face of the actor was either normally visible or it was occluded to exclude any information about the gazing behavior. The pointing movement was still completely visible. With the second factor, we manipulated the visibility of the target locations. The targets were either visible to the participant or they were digitally removed from the movies. In the first condition, both the gaze shift and the targets were visible. In the second condition, the gaze shift was visible, but the targets were missing. In the third condition, the face was occluded, but the targets were visible. And, in the fourth condition, the gaze shift was occluded and the targets were missing. It has to be noted, that the same movies were used in all four conditions.

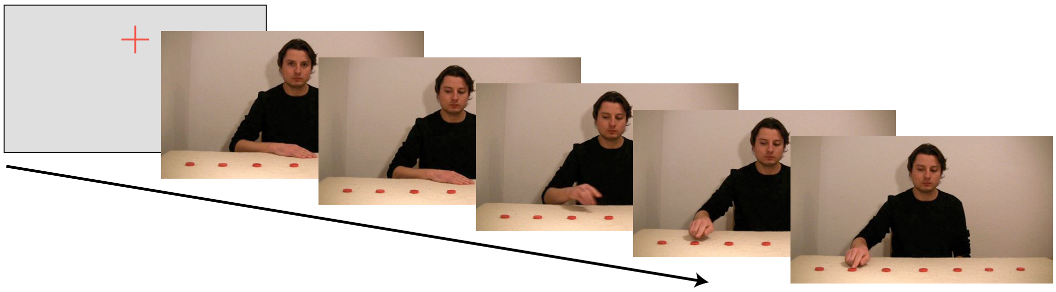

Figure 2: A sequence of frames representing one of the movies used as stimuli. Every trial started with a fixation cross and then the whole movie was presented. The actor in the movie gazed at one of the targets and executed a pointing movement towards the same target. Participants had to infer as soon as possible the relevant target and gaze on it.

Before the start of each movie a fixation cross was presented for 200 ms at the location were the face of the actor was going to appear (see Figure 2). The participants' task was to infer as soon as possible which target was going to be pointed by the actor. As soon as they made their decision, they had to gaze on the chosen target. In the case that they chose the wrong target, they were allowed to correct their choice. Therefore, depending on their accuracy they made one or more saccades to gaze on the correct target. Eye movements were measured from the start to the end of each trial.

Each participant was tested in two separate sessions, each lasting about 30 minutes. In total, each participant completed 240 trials (4 conditions $\times$ 6 targets $\times$ 10 repetitions). The order of trials was random and different for each participant. The experiment was preceded by a practice session to acquaint the participants with the task.

3.2.4 Data Analysis

The data analysis was focused on the number of saccades needed to land on the correct target and on the time it was needed to saccade on the correct target. We defined the start and end of a saccade when eye velocity exceeded or fell below a threshold of 30°/s and acceleration was above or below a threshold of $8000°/s^2$. Fixations were defined as the time between saccades. A correctly identified target was defined as the correct target on which participants kept their fixation for more than 200 ms.

3.3 Results

First, we focus our analysis on the number of saccades needed for target identification in the four different conditions and, second, on the timing of target identification in the four conditions.

3.3.1 Number of saccades for target identification

A repeated measures ANOVA with the main factors (a) presence of the gazing behavior, and (b) visibility of the targets was performed on the number of saccades needed to land on the correct target. The analysis revealed a significant effect of the presence of the gazing behavior ($F(1,5) = 38.111, P < 0.005$). In the conditions in which the participants saw the gazing behavior of the actor significantly less saccades were needed to identify the correct target in contrast to the conditions in which the gazing behavior was occluded (see Figure 3). On the other hand, no significant difference in the number of saccades was observed between the conditions in which the targets were either visible or occluded ($F(1,5) = 1.056, P = 0.351$). The interaction between the two factors did not reach significance ($F(1,5) = 1.354, P = 0.052$).

This analysis showed that a larger number of saccades was needed to land on the correct target, when the participants did not have the chance to see the actor's gazing behavior.

3.3.2 Timing of target identification

A repeated measures ANOVA with the main factors (a) presence of the gazing behavior, and (b) visibility of the targets was performed on the timing of target identification. The

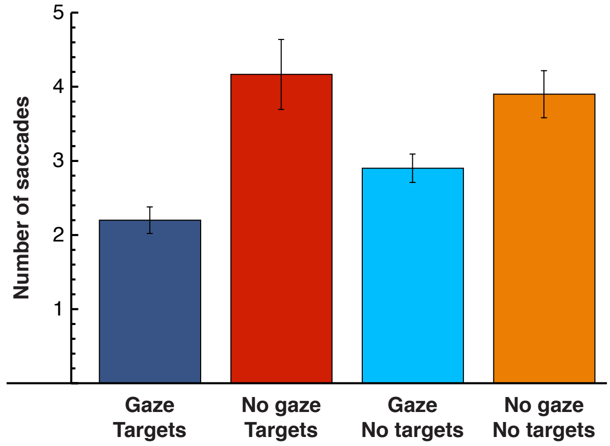

Figure 3: The number of saccades needed to land on the correct target object as a function of the different experimental conditions. The error bars denote standard errors of the mean.

timing of identification was calculated with respect to the point in time when the actor's finger touched the target. Negative values denote the cases in which the participants correctly landed on the target before the actor's finger actually touched it. On the other hand, positive values denote the cases in which the participants landed on the correct target after the actor's finger already touched it.

The repeated measures ANOVA showed a significant effect of the presence of the gazing behavior ($F(1,5) = 41.972, P < 0.001$). When the participants were allowed to see the actor's gazing behavior, they identified correctly the target 200-300 ms before the actor's finger touched the target. On the contrary, when the actor's gazing behavior was occluded, participants' saccades landed on the correct target simultaneously with the actor's finger (see Figure 4). No effect of the visibility of the targets was found ($F(1,5) = 3.616, P = 0.116$), and similarly no interaction between the two main factors ($F(1,5) = 3.243, P = 0.132$).

The availability of the actor's gazing behavior induced a consistent temporal advantage. The participants' saccades landed on the correct target relatively long before the actor's finger was positioned on the target.

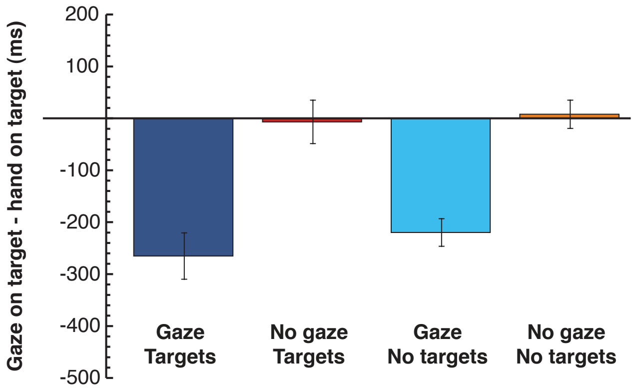

Figure 4: The timing of correct target identification as a function of the different experimental conditions. Negative values specify predictive gaze behaviors, i.e., the participants' gaze landed on the correct target before the actor's finger. The zero value specifies the case in which the participants' gaze landed on the correct target simultaneously with the actor's finger. The error bars denote standard errors of the mean.

3.4 Discussion

Eye movements often precede the motor behavior towards an object: we tend to first gaze at the object with which we intend to interact. Eye movements can thus reveal an individual's focus of attention and predict subsequent actions. Determining the direction of seen gaze toward a particular object in space is especially crucial in establishing joint attention. When an individual points to an object after having gazed at it, the availability of the cues indicating the chosen object dynamically changes over time. Initially, gaze direction is the only accessible cue, but as soon as the arm starts moving also the kinematic cues are available.

We investigated the humans' capability to direct their own gaze to a specific target in space that was gazed by an actor. The actor's action started with a gaze directed towards the specific target and was followed by a complete pointing movement towards the same target. The actor's behavior was thus mimicking a natural pointing behavior toward an object of limited extent. As in natural behavior, the eye movements towards the target always preceded the hand movement. The main focus of the present study was directed to the temporal characteristic of the humans' capability in directing their gaze. If the humans are able to predict the forthcoming relevant location indicated by the actor's gaze, then this capability should be reflected in gaze movements that anticipate the moment in time at which the actor's finger points on the specific target.

In fact, our results show that not only can humans gaze accurately on the target of interest that is gazed by the actor, but they also complete their gazing behavior before the actor's hand approaches the target. It is thus clear that predicting behaviors are utilized. This conclusion was supported by two interrelated results. Firstly, when the actor's gaze change was available, participants performed fewer saccades in order to land on the correct target than when the target identification was based on the hand movement only. Most of the times, the correct target was identified with two saccades: a first larger saccade and a second corrective saccade towards the center of the target. Secondly, when the actor's change was available, participants identified the correct target with a 200-300 ms temporal advantage in comparison with the conditions, in which the actor's gaze behavior was occluded. The actor's gaze triggered a rapid and accurate response on the target object. On the other hand, when gaze information was not available, participants' gaze still led the hand movements of the actor, but was comparatively slower in identifying the target objects. The second factor that was varied in our study, namely, the presence/absence of targets, had limited effects on the spatial accuracy of the participants' gazing behavior, but no impact on the timing of the responses.

These findings suggest that other's gaze direction is an essential predictive cue about the final location of a pointing movement. Observers thus activate action plans based on the actor's gaze direction and kinematic cues to produce proactive eye movements. This knowledge can be thus used by the robotic system to monitor the attention of a human cooperation partner and predict his/her actions. Moreover, since the predictive behaviors are intrinsic in the human behavior, an interaction with a robotic system that implements these predictive behaviors will be enriched by a more human-like experience.

4 Study II: Potentially purposeful actions divert overt attention

4.1 Introduction

The direction of an individual's gaze specifies which object the individual is focusing the attention to and a shift in gaze direction specifies a change of the object of attention. Gaze following is a key element in joint attention, because it drives the understanding of other individual's intentions and it constitutes the basis for the prediction of other individual's behavior. Under normal conditions people show a strong tendency to follow another's person gaze. Shifts in other's gaze direction lead to overt or covert changes in the behavior of an observer, that is, observer's gaze is directed toward the same target or the target is attended without the observer shifting the gaze toward it. However, overt or covert shifts of attention are not obligatory. In fact, in terms of parsimony and limited attentional resources, a system that reacts indiscriminately to every gaze direction change would be highly inefficient. Humans perform up to three eye movements per second and since most of these eye movements are most likely irrelevant to an observer, a fundamental question arises: How do humans distinguish the relevant shifts in other's gaze direction from the irrelevant ones? It is clear that processes should exist that prioritize potentially relevant gaze shifts to which the visual system automatically responds.

Joint attention can be measured using standard methods of attention research. For instance, in different variations of the spatial cueing tasks (Posner, 1980) it has been shown that participants covertly orient their attention in the direction to which the eyes of a centrally presented face gaze at, although this cue is spatially uninformative (Driver et al, 1999; Friesen and Kingstone, 1998). The cueing effect was expressed as a response time advantage in the detection of targets at the location that was gazed at in contrast to the detection of targets presented at the non gazed location. More interestingly, it has been found that gaze shifts can induce overt orienting responses (Crostella et al, 2009; Kuhn and Benson, 2007; Kuhn and Kingstone, 2009; Mansfield et al, 2003; Ricciardelli et al, 2002, 2009). In Ricciardelli et al (2002), participants were asked to make a speeded saccade to the left or right of fixation, as indicated by a central instruction stimulus. A centrally presented image of a face with eye gaze directed either toward the target location (congruent) or in the opposite direction (incongruent) acted as the distracter cue. Participant's saccade latencies were faster for congruent than for incongruent trials, mirroring the congruency effect in covert attention studies. Furthermore, participants performed unwanted saccades to the target at which the gaze was directed to, although the instruction cue required a saccade in the opposite direction. This effect was strongest when the distracter cue was presented slightly before the appearance of the instruction cue. Taken together, these studies evidence that observation of shifts in other's gaze direction can trigger overt reflexive-like responses. However, the percentage of saccades following the observed gaze instead of the instruction cue was relatively low, generally around 10%. This should not surprise us, since the gaze shifts might have been perceived as ordinary eye movements lacking any particular intention. We might hypothesize that the likelihood of overtly following other's gaze shifts should increase with an increased relevance of the observed eye movements.

In natural behaviors, a very tight link exists between gaze and hand movements that lead to the execution of actions. Eye movements briefly precede hand movements in a highly predictive manner (Johansson et al, 2001), and fixations are close to the site of action (Ballard et al, 1992). Because gaze typically remains at the contact location until around the time of action completion, the gaze direction can be a very accurate predictor of the gazer's region of imminent interest at least during the initial phase of a hand movement. Being able to lock on to these specific eye movements and disregard most of other irrelevant eye movements would certainly be beneficial for an efficient interaction between two individuals. This strong coupling between eye and hand movements could be used as a heuristic to distinguish relevant from irrelevant shifts in other's gaze direction.

In this study, we aim to exploit the gaze-hand coupling observable in natural behaviors to show that purposeful gaze shifts can have a stronger influence on gaze following than ordinary eye movements. For this purpose we implemented a modified version of the gaze cueing paradigm, in which we measured the directional errors of saccades and their latencies. Instead of presenting an image of a face with a diverted gaze as a distracter cue only, we also centrally presented a hand in the initial phase of a reaching movement towards one of the targets (see Figure 5). In practice, two distracter cues could be presented either separately or simultaneously. When presented separately, simple gaze or hand cueing are measured. On the other hand, under simultaneous presentation, the effect that hand movements have on gaze cueing is determined. It is clear that under simultaneous presentation the directions of gaze and hand can be manipulated independently. Both gaze and hand can be directed concurrently toward the same target (matched) or toward opposite targets (unmatched). When the distracter cues are matched, we have to expect inevitably that their cueing effect will be greater than when either of the two cues is presented separately. However, when the distracter cues are unmatched, several interesting questions arise. Do the cueing effects of gaze and hand cancel each other? Does one prevail over the other? If so, which one takes over? Is the magnitude of the cueing effect comparable with that obtained with a single distracter cue?

We hypothesized that the gaze shifts presented in combination with the hand movements should be interpreted as purposeful eye movements and, therefore, trigger stronger gaze cueing effects than simple gaze shifts. Moreover, the hand movements used in our study are only of small extent and, because they do not develop in a full reaching movement toward the target, they may provide partially ambiguous directional information. As a consequence, we expected that, perhaps counterintuitively, a gaze cueing effect should arise independently of the hand movement direction and it should comfortably exceed the simple gaze cueing effect.

4.2 Materials and Methods

4.2.1 Participants

Eight males and eight females aged from 19 to 32 year participated in the experiment after providing written informed consent. All had normal or corrected-to-normal vision and were naïve as to the purpose of the experiment. The experimental protocol was conducted according to the Declaration of Helsinki.

4.2.2 Apparatus

In the experiment, the participant sat 54 cm in front of a 20 inch computer monitor (Iiyama Vision Master Pro 514) with a vertical frequency of 100 Hz at a resolution of 1600×1200. The head was stabilized by a chin rest. For stimuli presentation and data collection we used MATLAB with Psychtoolbox (Brainard, 1997; Cornelissen et al, 2002; Pelli, 1997). Eye movements were monitored by the head-mounted Eyelink II system (SR Research Ltd., Mississauga, Ontario, Canada), which samples gaze positions with a frequency of 500 Hz. Viewing was binocular, but only the dominant eye was recorded. Calibration was performed at the beginning of the recorded session and drift correction was performed every 10 trials to ensure optimal accuracy.

4.2.3 Stimuli and Procedure

Stimuli were colored digital pictures (1200×900 pixels, visual angle: $31.64° \times 23.99°$) of a young woman's upper body comprising the face and the hand that was laid on the table in front. Three red objects ($2.34° \times 0.64°$) were also present in the picture, one on either side of the table and one in it's center. The left and right objects served as saccade targets. In the neutral stimulus, the gaze of the woman was directed towards the middle object and the hand was also laid on the same object. The distracter stimuli were the same as the neutral stimulus except for a change in the gaze direction and/or hand position (see Figure 5). The distracter stimuli were of four categories: (a) gaze-direction change only, (b) hand-direction change only, (c) simultaneous and

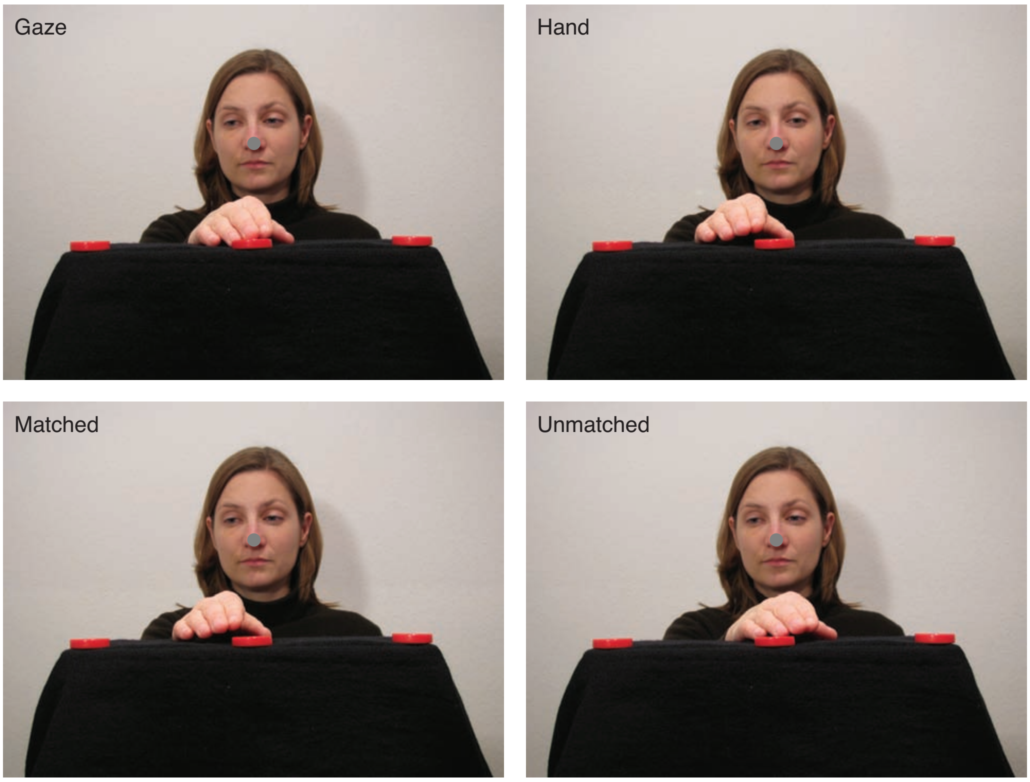

Figure 5: Distracter stimuli. Top left, the gaze is directed to the left object (gaze condition). Top right, the hand is moved in the direction of the left object (hand condition). Bottom left, the gaze is directed to the left object and the hand is moved in the direction of the same object (matched condition). Bottom right, the gaze is directed to the left object and the hand is moved to the object on the opposite side (unmatched condition). Participants fixated the gray dot that changed the color to either black or white and by this instructed the participants to saccade to the left or right object, respectively.

matched gaze- and hand-direction change, (d) simultaneous but unmatched gaze- and hand-direction change. An apparent motion of the gaze and/or hand change was induced by presenting successively the neutral and one of the distracter stimuli. We will refer to these conditions as gaze, hand, matched and unmatched conditions.

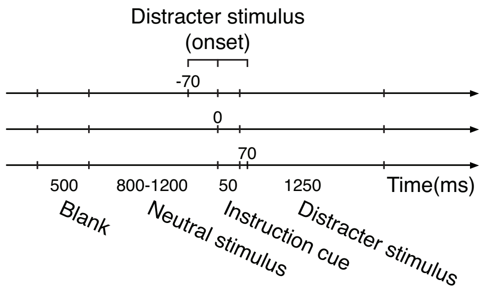

Figure 6: A timeline illustrating the course of each trial. After a blank screen was presented for 500 ms, the neutral stimulus was shown. In the neutral stimulus, the gaze was directed on the centrally located object and the hand was laid on it. A medium gray fixation dot was located on the nose. After a variable period between 800 and 1200 ms, the instruction cue occurred, i.e., the fixation dot changed to either black or white instructing the participants to perform a saccade to either the left or the right object, respectively. The instruction cue was presented for 50 ms. The distracter stimulus was presented relative to the instruction cue onset (SOA: -70, 0, 70 ms). A blank screen replaced the distracter stimulus 1250 ms after the instruction cue offset.

It has to be noted that the stimuli were constructed by cropping and then joining the upper and lower parts of a set of images. In this way, the part of the stimulus representing a specific gaze or hand direction was identical across all the stimuli that required the specific gaze or hand direction. For instance, in Figure 5 the upper part of the stimuli, in which the gaze is directed to the left, is identical in the top left, bottom left and bottom right stimuli. The same holds for other combinations of gaze and hand directions.

The timeline of each trial is illustrated in Figure 6. Each trial began with the presentation of the neutral stimulus. A medium gray fixation dot ($0.5°$) was located on the nose. The fixation dot was located slightly closer to the eyes than to the hand to compensate for the relative size difference between the distracter cues. A color change of the fixation dot (either to black or to white) instructed the participant to perform the saccade to the left or to the right object, respectively. This instruction cue occurred between 800 and 1200 ms after the presentation of the neutral stimulus. The instruction cue was presented for 50 ms. The distracter stimulus containing the change in gaze and/or hand direction was presented at different time intervals relative to the presentation of the instruction cue (stimulus onset asynchronies, SOA: -70, 0, 70 ms). The instruction cue and the distracter stimulus were presented simultaneously when the SOA was equal to zero. The distracter stimulus preceded or followed the instruction cue when the SOA were negative or positive, respectively. The distracter stimulus was replaced by a blank screen that introduced the next trial 1250 ms after the instruction cue offset.

Each participant was tested in four separate sessions, each lasting about 20 minutes. In total, each participant completed 960 trials (4 distracter conditions $\times$ 2 distracter directions $\times$ 3 SOAs $\times$ 2 saccade instruction cues $\times$ 20 repetitions). The order of trials was random and different for each participant. Participant were instructed to perform a saccade to one of the two objects. The direction of the saccade was determined by the color of the instruction cue. Moreover, the participants were instructed to disregard any directional information given by the distracter stimuli. The experiment was preceded by a practice session to let the participants learn to associate a particular color of the instruction cue with the specified direction of the saccade.

In summary, four main distracter conditions were measured each with the three different SOAs. In the gaze condition, the gaze was directed towards one of the two targets. In the hand condition, the hand moved in the direction of one of the two targets. The extent of the movement was about one quarter of the total distance between the central and one of the external targets. In the matched condition, the gaze was directed towards one of the two targets and the hand moved in the direction of the same target. In the unmatched condition, the gaze was directed towards one of the two targets, however, the hand moved in the opposite direction towards the other target.

4.2.4 Data Analysis

We defined the start and end of a saccade when eye velocity exceeded or fell below a threshold of $30°/s$ and acceleration was above or below a threshold of $8000°/s^2$. We analyzed the directional errors by focusing on the first saccade that followed the instruction cue and had an amplitude larger than $2°$. In addition, we applied a threshold based on the latency of the saccade to eliminate anticipatory saccades (below 100 ms) and delayed saccades (above 500 ms). Any trials in which the eye-tracker loss occurred were excluded from the analysis. In total, 5.6% of trials were excluded.

Directional errors were computed by calculating the proportion of erroneous saccades, i.e. saccades made in the direction opposite to that indicated by the instruction cue. To analyze saccade latencies we considered the correct trials only, i.e. trials in which the saccade was correctly made in the direction indicated by the instruction cue and landed in the 2° area around the center of the specified target. Congruent trials in the gaze condition and in the hand conditions were the trials in which the instruction cue required a saccade in the same direction as the gaze or the hand, respectively. Given that in the matched and unmatched conditions two cues were always presented simultaneously, we chose to define the congruent trials with respect to the direction specified by the gaze cue.

There were no significant differences in the directional errors and in the saccade latencies between trials in which the instruction cue required to saccade to the left or to the right, so in all further analyses we collapsed these conditions.

4.3 Results

4.3.1 Influence of congruency and SOA

Our first interest was focused on the proportion of eye movements made in the direction contrary to that indicated by the instruction cue. We analyzed these directional errors with respect to the congruency of the distracter cues and the time interval between the presentations of the distracter cue and the instruction cue (SOA). We ran a repeated measures ANOVA with congruency and SOA as main factors for each condition separately. Figure 7, left column, shows the proportion of directional errors for each condition.

In the gaze condition, we found a significant effect of the main factors of congruency ($F(1, 15) = 19.693, P < 0.001$) and SOA ($F(2, 30) = 5.216, P < 0.05$). The interaction between these factors was also significant ($F(2, 30) = 5.066, P < 0.05$). In the hand condition, we found a significant effect of both congruency ($F(1, 15) = 12.397, P < 0.01$) and SOA ($F(2, 30) = 3.98, P < 0.05$), but no interaction effect ($F(2, 30) = 0.261, P = 0.772$). In the matched change condition, the analysis revealed a significant effect of both congruency ($F(1, 15) = 53.737, P < 0.001$) and SOA ($F(2, 30) = 13.087, P < 0.001$). The interaction was also significant ($F(2, 30) = 9.786, P < 0.001$). In the unmatched condition, we found a significant effect of both congruency ($F(1, 15) = 31.122, P < 0.001$) and SOA ($F(2, 30) = 24.044, P < 0.001$). The congruency by SOA interaction was also significant ($F(1, 15) = 42.401, P < 0.001$).

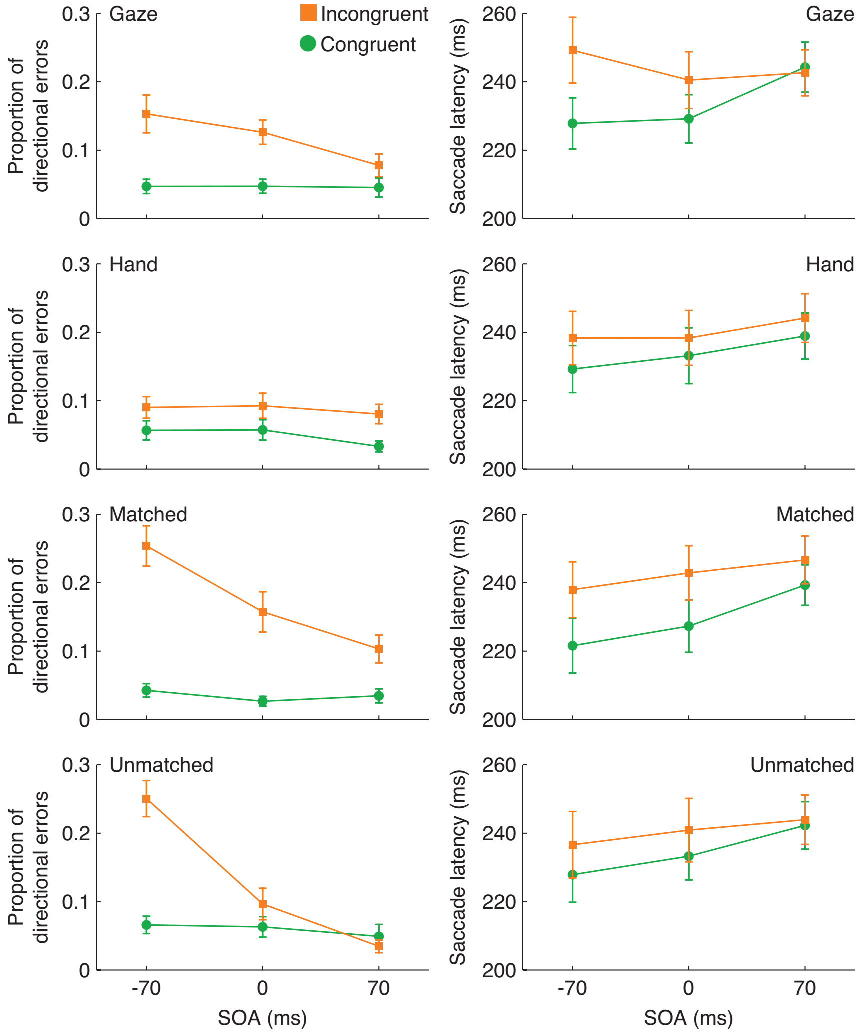

Figure 7: Left column, mean proportion of directional errors as a function of SOA made for congruent and incongruent trials for the four distracter conditions. Right column, mean saccade latencies as a function of SOA for congruent and incongruent trials for the four distracter conditions. Error bars denote standard errors of the mean.

These analyses showed that in all four distracter conditions (a) the directional errors were more frequent in the incongruent than in the congruent trial type, and (b) the directional errors varied as a function of SOA. The negative SOA induced the highest proportion of directional errors which than linearly declined for the zero and for the positive SOAs. The interaction revealed a different pattern in the proportion of directional errors as a function of SOA for the incongruent and the congruent trail types in all conditions except the hand condition. The decrease in the proportion of directional errors from negative to positive SOAs was limited to the incongruent trial types.

We ran the same analyses on the saccade latencies in the correct trials (Figure 7, right column). In the gaze condition, we found a significant effect of congruency ($F(1, 15) = 16.382, P < 0.001$) and SOA ($F(2, 30) = 11.287, P < 0.001$). The interaction was also significant ($F(2, 30) = 22.012, P < 0.001$). In the hand condition, the analysis revealed a significant effect of congruency ($F(1, 15) = 30.174, P < 0.001$) and SOA ($F(2, 30) = 5.777, P < 0.01$), but no interaction effect ($F(2, 30) = 0.822, P = 0.449$). In the matched condition, there was a significant effect of congruency ($F(1, 15) = 84.578, P < 0.001$) and SOA ($F(2, 30) = 11.901, P < 0.001$), and a significant interaction between these two factors ($F(2, 30) = 6.099, P < 0.01$). In the unmatched condition, we found a significant effect of congruency ($F(1, 15) = 6.968, P < 0.05$) and SOA ($F(2, 30) = 9.858, P < 0.001$), but no interaction effect ($F(2, 30) = 1.066, P = 0.357$).

The analyses on saccade latencies showed that in all four distracter conditions responses (a) were slower for incongruent than for congruent trial types, and (b) were increasingly slower from negative to positive SOAs. The interaction effects, when significant, reflected a larger difference between incongruent and congruent trial types at the negative SOA than at zero or positive SOAs.

4.3.2 Congruency effects between conditions

Let us now consider the directional error congruency effect between distracter conditions. We computed the size of the congruency effect by subtracting the proportion of directional errors in the congruent condition from the proportion of directional errors in the incongruent condition. The directional error congruency effects at each SOA are shown in Figure 8. A separate repeated measures ANOVA with distracter condition type as within-subjects variable was conducted for each SOA.

The directional error congruency effect was significantly different between the four distracter conditions, when the distracter stimuli were presented before the instruction cue

(SOA $= -70$ ms: $F(3,45) = 14.527, P < 0.001$), simultaneously with the instruction cue (SOA $= 0$ ms: $F(3,45) = 7.606, P < 0.001$), and after the instruction cue (SOA $=$ 70 ms: $F(3,45) = 7.807, P < 0.001$). Subsequent pairwise comparisons with Bonferroni corrections revealed the following differences between distracter conditions. When SOA was negative, the congruency effects in both the gaze condition (0.106 $\pm$ 0.03 SEM) and the hand condition (0.033 $\pm$ 0.016 SEM) were significantly lower than in both the matched (0.211 $\pm$ 0.03 SEM) and unmatched conditions (0.185 $\pm$ 0.021 SEM). When the SOA was zero, only the congruency effect in the matched condition (0.131 $\pm$ 0.029 SEM) was significantly higher than in the hand condition (0.035 $\pm$ 0.021 SEM) and in the unmatched condition (0.033 $\pm$ 0.018 SEM). The gaze condition (0.079 $\pm$ 0.018 SEM) did not differ significantly from any other condition. When the SOA was positive, only the congruency effect in the unmatched condition (-0.014 ± 0.014 SEM) was significantly lower than in the gaze condition (0.032 $\pm$ 0.012 SEM), in the hand condition (0.047 $\pm$ 0.011 SEM), and in the matched condition (0.068 $\pm$ 0.018 SEM).

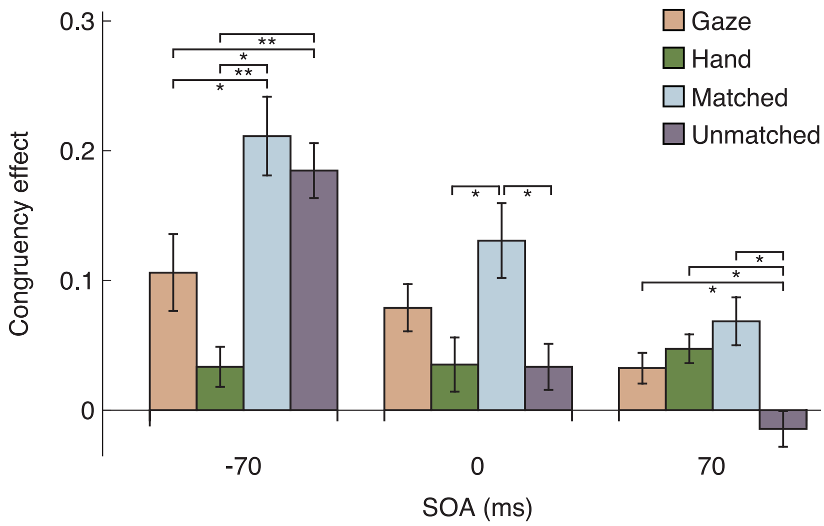

Figure 8: Directional error congruency effects of the four distracter conditions grouped for the different SOAs. The congruency effect is calculated by subtracting the proportion of directional errors in the congruent condition from the proportion of directional errors in the incongruent condition. Error bars denote standard errors of the mean. Line with asterisks indicate significant differences between conditions revealed by Bonferroni corrected t-tests (*: $P < 0.05$, **: $P < 0.001$).

The congruency effect was strongest when gaze and hand cues were presented simultaneously (matched or unmatched conditions) and before the instruction cue (SOA = -70 ms). This means that the largest amount of saccades made in the direction of gaze and contrary to that indicated by the instruction cue was induced by the concurrent presence of the hand movement, but it was independent of its direction. This effect then diminished at zero and positive SOAs to the level of the congruency effects of the conditions in which the gaze or hand cues were presented in isolation. Only the matched condition still showed a larger congruency effect at zero SOA.

In addition, we computed the difference between saccade latencies by subtracting the responses to the congruent trials from the incongruent trials and ran the same analyses as previously on the directional error congruency effect. Although these analyses showed a significant effect between conditions at negative SOA ($F(3, 45) = 4.159, P < 0.05$) and positive SOA ($F(3, 45) = 3.318, P < 0.05$), but not at zero SOA ($F(3, 45) = 2.448, P = 0.076$), no subsequent pairwise comparison did reach significance. It follows that the differences in the congruency effect between distracter conditions were limited to the directional errors.

4.4 Discussion

Humans have a natural tendency to follow sudden changes in other's gaze direction, because they can direct attention to important aspects of the environment and they can help sustaining the interaction between individuals. This tendency was shown experimentally among other methods with a gaze cueing task in which a centrally presented image of a face with diverted gaze was used as a distracter cue and the participants' task was to saccade in the direction indicated by an instruction cue (e.g., Ricciardelli et al, 2002). Although the diverted gaze was uninformative about the direction in which the saccade had to be performed, participants exhibited a propensity to saccade in the same direction as the observed gaze. About 10% of the saccades followed the gaze shift instead of the direction indicated by the instruction cue. A reason why this percentage was relatively low might be attributed to the fact that the observed gaze shifts were actually lacking any particular intention. Therefore, we hypothesized that other's gaze shifts that express the potentiality of a purposeful action should be more difficult to disregard and thus induce a larger percentage of unwanted saccades in the gazed at direction.

It is known that in natural behavior eye and hand movements are tightly linked and gaze typically lands on an object before the hand reaches it. This implies that the gazed-at locus can accurately indicate the position at which a forthcoming action is going to be executed. In the present study, we wanted to exploit this coupling between gaze and hand movements to create distracter stimuli that convey the potentiality of a purposeful action. We measured the tendency to perform unwanted saccades in the direction of the distracting stimulus in four different conditions. In the first two conditions, we used either a gaze shift or a small hand movement as distracter stimuli. In the other two conditions, we combined the gaze shifts and the small hand movements in a single distracter stimulus. These last two conditions differed with respect to the matched or unmatched direction of the single distracters cues. In one case, both the gaze shift and the hand movement were directed toward the same target, whereas in the other case, the gaze shift and the hand movement were directed toward the opposite targets. All conditions were measured with different stimulus onset asynchronies (SOA) varying between -70 ms to 70 ms.

First of all, we present the effects that the distracter cues induced in the different conditions. We analyzed both the saccadic directional errors and the saccade latencies. The directional errors correspond to the proportion of saccades made in the direction contrary to that indicated by the instruction cue, and are thus computed on the whole set of data. On the other hand, the saccade latencies are restricted to the saccades of the correct trials only. Both the directional errors and the saccade latencies were analyzed with respect to the congruency between the distracter direction and the instruction cue direction, and with respect to the SOA.

In the single distracter cue conditions, we found that both the gaze shift and the hand movement had an effect on the directional errors. When the distracter cue was incongruent with the instruction cue, a larger proportion of unwanted saccades was triggered in the direction of the distracter cue than when both the distracter and the instruction cue pointed in the same direction. However, the effect was more marked with the gaze shift cue than with the hand movement cue. The effect on directional errors was modulated by the different SOAs. The gaze shift induced a larger effect with a negative SOA, that is, when the distracter cue was presented before the instruction cue, and than diminished at zero and positive SOAs. On the contrary, the hand movement induced a stable effect at all SOAs. The saccade latencies analysis showed that the responses were slower for incongruent than for congruent trials, and that the responses were increasingly slower from negative to positive SOAs. Taken together, the findings with the gaze shift distracter cue replicated the magnitude of the effect previously found in gaze cueing tasks

(Crostella et al, 2009; Kuhn and Benson, 2007; Kuhn and Kingstone, 2009; Mansfield et al, 2003; Ricciardelli et al, 2002, 2009). In addition, we showed that also small hand movements can induced very similar effects.

The conditions in which the gaze shifts and the hand movements were jointly used as distracter cues were the critical conditions of the present study. In the matched condition, when both the gaze shift and the hand movement were directed toward the same target, we found a pattern of results that resembled the effects measured in the gaze shift only condition. It is, however, noticeable that the amount of unwanted saccades in the direction of the distracter cues was strongly enhanced especially when the distracters were presented before the instruction cue (SOA: −70 ms). On the other hand, the saccade latencies did not change with respect to the gaze only condition. More interesting were the effects in the unmatched condition, in which the gaze shift and the hand movement were directed towards the opposite targets. In principle, the interaction between the two distracter cues could lead to very different effects, ranging from mutual inhibition to single distracter cue enhancement. In fact, we found that the gaze shift influence was enhanced by the concurring hand movement, although the hand moved in the opposite direction. Similarly to the matched condition, this effect was very pronounced when the distracters were presented before the instruction cue (SOA: −70 ms). At zero and positive SOAs, the effect vanished almost completely suggesting a mutual inhibition between the two distracter cues. The effects on saccade latencies in the unmatched condition duplicated those found in the gaze only condition and in the matched condition.

As a second step, since the effects of gaze shift and hand movement were modulated by the time interval between the presentation of the distracter cue and the instruction cue, we compared the conditions separately for each SOA. The direct comparison between conditions showed a difference between the four distracter conditions in the proportions of directional errors, but no difference between saccade latencies. More specifically, both the matched and the unmatched conditions induced a larger proportion of directional errors than both the gaze and hand conditions, when the distracter stimuli were presented before the instruction cue (SOA: −70 ms). At zero SOA, only the matched condition was still showing an increased proportion of directional errors, which vanished when the distracter stimuli were presented after the instruction cue (SOA: 70 ms). In this last case, the proportion of directional errors in the unmatched condition dropped drastically and thus differed from the other conditions that still preserved an effect of congruency.

The fact that the congruency effect in the matched condition was larger than in the gaze and hand conditions can be simply explained by the fact that the simultaneous presentation of the gaze shift and the hand movement directed toward the same target contributed to an addition of the single cueing effects. However, the effects found in the unmatched condition need a different explanation. The gaze shift and the hand movement were directed to the opposite targets, therefore, both cues were actually put into conflict with each other. Our data unveiled an intricate behavior between the two cues that was modulated by the timing of the presentation of the distracter cue with respect to the instruction cue. At negative SOA, the gaze cue predominated over the hand movement cue and was even enhanced by it showing a comparable magnitude to the effect obtained in the matched condition. At zero SOA, the gaze cue was still producing a congruency effect, although reduced. On the other hand, at positive SOA the gaze directional cue lost almost completely its cueing effect. This temporal modulation might be ascribed to the different temporal modulations observable in the single cue conditions: the congruency effect in the gaze condition diminished as a function of SOA, whereas the congruency effect in the hand condition stayed stable and did even slightly increase.

The increased proportion of unwanted saccades in the direction of the gaze shift, when the hand movement was simultaneously presented, is a particularly surprising finding given that the effect was independent from the direction of the hand movement. The presence of the hand movement strongly decreased the ability to inhibit the diverting effect of the gaze shift. Since, in natural behaviors, eye and hand movements are tightly linked just before action execution, this failure in inhibiting gaze-following might actually help and support the interaction with other humans. If humans' overt attention is automatically directed toward a specific target in space, whenever the interacting partner gazes at and starts performing an arm movement towards this same target, there will always be an advantage in predicting the behavior of the interacting partner. Therefore, we might hypothesize that the mere presence of a sudden hand movement might have been interpreted as a sufficient indication of a forthcoming relevant action that consequently enhanced the saliency of the directional cue provided by the gaze. Processes should thus exist that prioritize potentially relevant actions to which the visual system automatically responds.

Our findings may also have some implications in the design of human-robot interfaces based on the monitoring of eye movements for the prediction of human's actions. Implementing a heuristic to discriminate relevant from irrelevant eye movements would be certainly beneficial in reducing the amount of information to which the robotic system should respond or adapt its behavior.

In sum, we showed that the visual system has a strong tendency to follow the gaze direction of an external observer, when the external observer simultaneously executes gaze and hand movements. Most importantly, this tendency was independent of the hand movement direction and way larger then when the gaze shifts were presented in isolation. Humans thus take advantage of the coupling of eye and hand movements in natural behaviors to identify and predict as quickly as possible the regions of interest of the interacting partner.

5 Study III: Gaze behavior in cooperative action

5.1 Introduction

Only a small fraction of the information in a single gaze can be attended and retained in working memory. Thus, there must be mechanisms that guide this selection process. Deployment of gaze is an overt manifestation of this allocation of attention. When humans explore the environment they do not only look at it, but they also interact with it by manipulating objects that are part of the environment. A central question is thus how we use our eyes to obtain the information we need for action. An even more intriguing question concerns the situation in which two humans interact in a common task by sharing a common space and, most of the times, common goals. The purpose of the present study was to measure the gaze behavior of two cooperating partners during an interaction involving a simple task based on grasping and moving an object to reach a common goal.

Johansson et al (2001) showed that gaze supports hand movement planning by marking key positions to which hand action is subsequently directed. The salience of these gaze targets did arise from functional sensorimotor requirements of the task. In their study, participants had to grasp an object and move it around an obstacle to press a target-switch located above the obstacle. Participants never fixated or tracked their own hand or their own moving object during the task. They found that gaze and hand movements were linked with respect to key landmarks with gaze always leading the hand. Gaze was almost exclusively directed to object involved in the task and certain key landmarks were obligatory gaze targets.

In the present study, we used a very similar task. The difference was that each of the two participants had to move his/her own object around an obstacle and then, instead of pressing a target-switch, they had to make contact with the partner's object. In this variation of the task, the target location varied in space across trials and had to be jointly determined between the two interacting partners. In this case, the predictability of the contact location was considered to be normal, since both participants were in charge of adjusting their own actions. In a second variation of the task, we instructed one of the two participants to disregard completely the hand movements of the other participant. The direct effect was that the first participant determined the contact location and the second participant had to adjust his/her own action to make successfully contact with the first participant's object. In this second case, the predictability of the contact location was considered to be low, since the second participant did not have any control about the location where the two objects had to make contact.

5.2 Materials and Methods

5.2.1 Participants

Fourteen participants (six male and eight female) aged from 19 to 29 years took part in the experiment after providing written informed consent. All had normal or corrected-to-normal vision and were naïve as to the purpose of the experiment. The experimental protocol was conducted according to the Declaration of Helsinki.

5.2.2 Apparatus

In the experiment, each pair of participants sat behind a table facing each other at a 120 cm distance. The head of both participants was stabilized by a chin rest. Eye movements of each participant were monitored by a head-mounted ViewPoint eye tracker system (Arrington Research Inc., Scottsdale, AZ), which samples gaze positions with a frequency of 60 Hz. Viewing was binocular, but since all the relevant actions occurred in the fronto-parallel plane only the dominant eye was recorded. Calibration was performed at the beginning of the recorded session and constantly monitored. If the accuracy of the eye movements recording degraded during a recording session, the system was re-calibrated.

Each of the two eye tracker systems had a head-mounted 30 Hz video-camera that was directed towards the peripersonal space in front of the participant. In this way, eye movements were superimposed offline on the video recordings to monitor and study the interaction of gaze behavior, hand movements and object handling.

5.2.3 Procedure

At mid-distance between the two participants a structure was positioned that served as reference for the hand movements that the participants had to perform. The cross-shaped structure was 25 cm wide and 20 cm high (see Figure 9). Each of the two participants held a 10 cm wide, 5 cm deep and 2 cm high object. The bottom of the structure served as the starting position for the task at hand. Both participants had to grasp and move the object around the obstacle (the wide horizontal bar of the structure) and make contact with the object of the other participant in the area above the set-up. Each participant moved his/her own object on one side of the structure. After contact was made the objects had to be moved back to the starting position along the same trajectory. Each movement from the starting position until contact was made with the partner's object was defined as one trial measurement (see Figure 10).

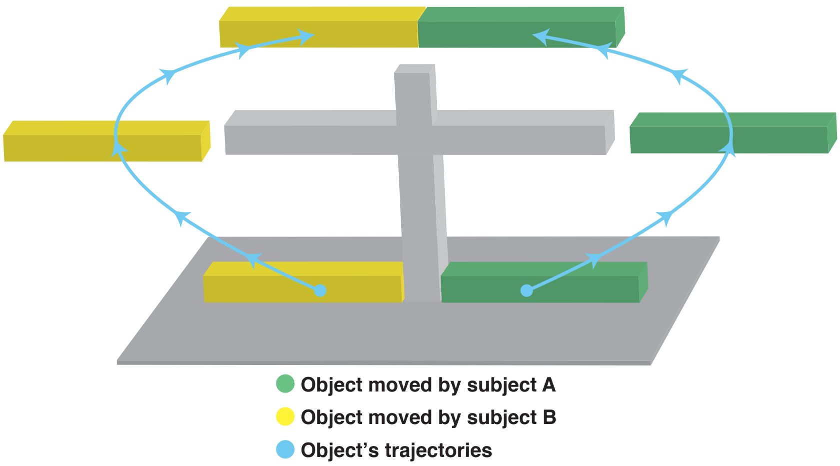

Figure 9: Sketch of the set-up. The gray structure served as a reference for the hand movements participants had to perform. Each of the participants held an object (yellow or green) that had to be moved along the cyan trajectory to make contact above the structure with the object of the other participant.

The experiment consisted of two experimental conditions. In the first condition (normal predictability), the task was to make contact with the object of the opposing partner. To this end, both participants had to adapt their movements to the movements of the other participant to complete successfully each trial. In the second condition (low predictability), one of the participants had the instruction to execute the same movement as in the previously described condition, but the movement did not have to be adjusted to the trajectory of the partner's movement. Therefore, the partner had to fully adapt his/her own movement to the trajectory of the first participant. In all other respects, the two conditions were identical.

Each pair of participants was tested in both the normal predictability and in the low predictability conditions. The low predictability condition was measured twice in order to exchange the roles between the partners. Each partner did thus participate as the active and as the passive partner in the low predictability condition. The order of conditions was counterbalanced across participants' pairs. The experiment was preceded by some practice trials to accustom the participants with the task.

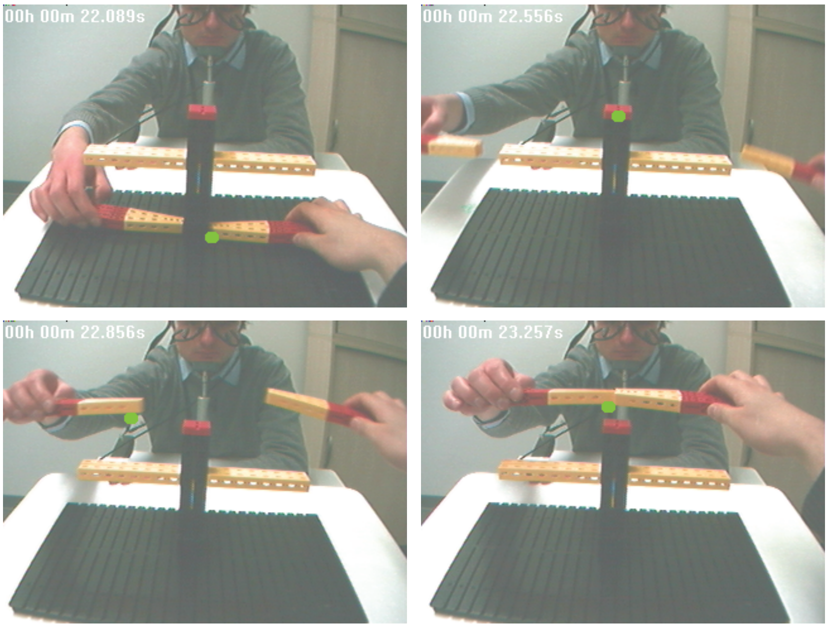

Figure 10: A sequence of frames (from top left to bottom right) taken by the head-mounted video-camera. One trial measurement is represented starting with both objects positioned on the bottom part of the structure and then moved around the obstacle. The trial was ended when the participants made contact with the two objects above the structure. The green dot in all the frames represents the gazing direction of the participant wearing the head-mounted video-camera. The same recordings were simultaneously obtained from the viewpoint of the other participant.

5.2.4 Data Analysis

Eye movement recordings of each participant were superimposed on the recordings taken with the head-mounted video-camera. With this procedure it was possible to define at each moment in time where the participant was directing his/her own gaze, where was his/her hand holding the object and where the partner's object was.

In sum, 2240 trials were recorded. All trials in which an eye recording loss or other problems occurred (e.g., the object slipped from the hand and fell) were excluded from further analysis. It has to be noted that if a trial was excluded from one participant's set of recordings, the same trial was also excluded from the set of recordings of the partner participant. In total, 13.2% of trials were excluded from analysis.

Individual eye movements data were analyzed. Using the video images from the head-mounted camera, the precise start and end of each trial was identified manually with an uncertainty of ± 16.7 ms (corresponding to the 30 Hz temporal resolution of the camera). The start of each trial was defined as the moment in time, when the object held by the participant started moving. The end of each trial was defined as the moment in time, when the object made contact with the partner's object. We then determined where and when did the participants look during the trial. We identified the following pattern: participants started looking at the tip of their object, soon afterwards they gazed on the top of the structure and before making contact with the partner's object they gazed on the tip of the partner's object. After this last saccade they followed the object until contact was made. In a very limited amount of trials an additional saccade was made towards the tip of the obstacle and it occurred before directing the gaze towards the top of the structure. This sequence of gaze direction changes is represented in Figure 11. The most relevant part of the gaze sequence was represented by the saccade towards the partner's object, therefore, the analysis was principally focused towards the frequency and timing of these object directed saccades. Starts and ends of the object directed saccades were extracted from the video images. The timing of the object directed saccades was calculated with respect to the contact time. The object directed saccades timing thus indicate how long before the contact time the saccades towards the partner's object were initiated.

5.3 Results

We evaluated the gaze direction changes with respect to two questions: First, where do humans normally look when they interact in a common task together with another human? Second, how does the gazing behavior depend on the cooperativeness of the other human?

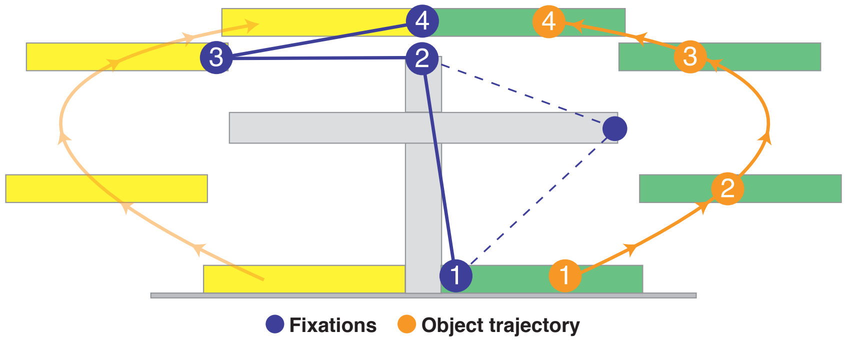

Figure 11: Stereotypical gaze behavior. The blue line represents the eye movements path. Participants started gazing at their own object (1, blue) and as soon as they started moving it, they directed their gaze toward the top part of the structure (2, blue). In few cases, an additional saccade was directed from the tip of their own object toward the tip of the obstacle. Before making contact with the partner's object they directed their gaze toward the partner's object – object directed saccade (3, blue). After this saccade they followed the partner's object until contact was made (4, blue). The orange line represents the trajectory of the object. The orange numbers indicate where the object was located when each of the gaze direction changes (blue numbers) occurred.

5.3.1 Stereotypical gaze behavior

From eye movements observation in the normal predictability condition a very stereotypical gaze behavior emerged: (1) at the start of each trial a fixation was directed towards the own object; (2) fixation was kept on a central location of the setup; (3) saccades were then regularly directed towards the partners object in the terminal phase of the movement prior to the contact between objects; (4) the gaze followed the object until contact was made (see Figure 11). In a very limited set of trials (5.3%) an additional saccade was directed toward the tip of the obstacle. The crucial aspect of this sequence was the object directed saccade, i.e. the saccade directed toward the partner's object.

This specific saccade was made in 91.2% of the trials, it was thus obligatory for an optimal execution of the task. In fact, since both partner's in the normal predictability condition executed regularly these object directed saccades, each partner was monitoring the behavior of the other partner in the last phase of the hand movement prior to the contact between objects. A closed feedback loop was thus established between the two partners: the hand movements of participant A were adjusted as a function of the hand movements of participant B, and, simultaneously, the hand movements of participant B were adjusted as a function of the hand movements of participant A.

In the low predictability condition, different gaze behavior patterns were observed. The participant that had to fully adapt his/her own hand movements for an optimal execution of the task showed exactly the same gaze behavior as in the normal predictability condition. The only difference concerned the timing of object directed saccades (see next section). On the other hand, the participant that was instructed to execute the hand movement without adapting his/her own trajectory showed a completely different pattern of eye movements. In essence, the gaze behavior was reduced drastically. Typically, the participants gazed on the tip of their own object and as soon as they started moving it, they gazed above the structure and maintained their gaze on that location until contact was made. This last saccade actually anticipated the location of contact between the two objects. No eye movements were directed towards the partner's object.

5.3.2 Timing of object directed saccades: comparison between conditions

In order to compare the normal predictability and the low predictability conditions we focused on the timing of object directed saccades. Usually, the moment in time at which a saccade is directed towards a specific location can be used as an indicator of the relevance of that location in the execution of a specific task at that specific moment in time. Since no object directed saccades were observed when the participants did not have to adapt their own hand movements to the partner's hand movements, these measurements were not included in the present comparison. Here, we compare the normal predictability condition data with the data of the participants that had to actively adapt their hand movements in the low predictability condition.

In order to compare the object directed saccade timing between the conditions we had to exclude that other differences between the conditions, such as a systematic difference in trial duration, might be present in our data. Figure 12, left panel, represents the durations of all trials in the normal and low predictability conditions. The overlap of

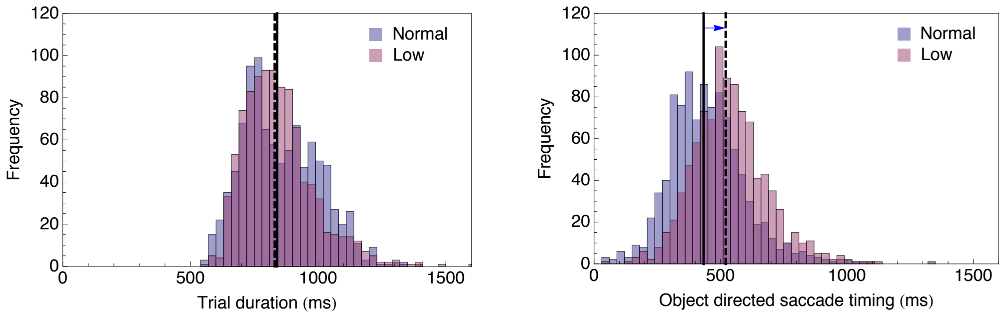

Figure 12: Left panel, durations of all trials in the normal (blue) and low (purple) predictability conditions. Right panel, objected directed saccade timing. In both panels, the vertical lines represent the mean value of each condition (solid line: normal predictability condition, dashed line: low predictability condition).

the two distributions indicates that the trial durations in the two conditions were fully comparable. The mean trial duration in the normal predictability condition was 839 ms (SD: 149 ms), whereas in the low predictability condition was 830 ms (SD: 138 ms). As a comparison, the right panel of Figure 12 shows the distributions of the object directed saccade timings in the two conditions. The overall shape of the distributions did not change: the standard deviation in the normal predictability condition was 152 ms and 146 ms in the low predictability condition. However, the two distributions differ by a shift along the temporal axis. The timing of the object directed saccades was calculated with respect to the contact time between the two objects, therefore, it expresses how long before the contact a saccade was initiated towards the partner's object. The shift of the whole distribution of the low predictability condition thus indicates an earlier initiation of the saccades toward the partner's object.

After grouping the data of each participant, the trial durations between the two conditions were compared with a paired t-test that confirmed that the durations were not significantly different ($t(13) = 0.252, P > 0.05$). Given that the trial durations in the two conditions were very similar, we proceeded with the comparison of the object directed saccades timing. The difference in the object directed saccade timing between the normal and the low predictability conditions is represented in Figure 13. A paired t-test showed that the object directed saccades started significantly earlier in the low predictability condition than in the normal predictability condition ($t(13) = 2.505, P < 0.05$). The object directed saccades were initiated on average 529 ms (SD: 76 ms) before contact in the low predictability condition, and 446 ms (SD: 96 ms) before contact in the normal predictability condition. The difference is thus quite substantial considering that the whole trial, from the start of the object movement to the contact with the partner's object, lasted on average slightly more than 800 ms. On average, thus, participants performed the object directed saccade shortly after the start of object's movement (417 ms after the start in the normal predictability condition and 323 ms after the start in the low predictability condition). The necessity of gazing on the partner's object after this very short time lapse prevented the execution of any other gazing behavior toward other objects in the environment.

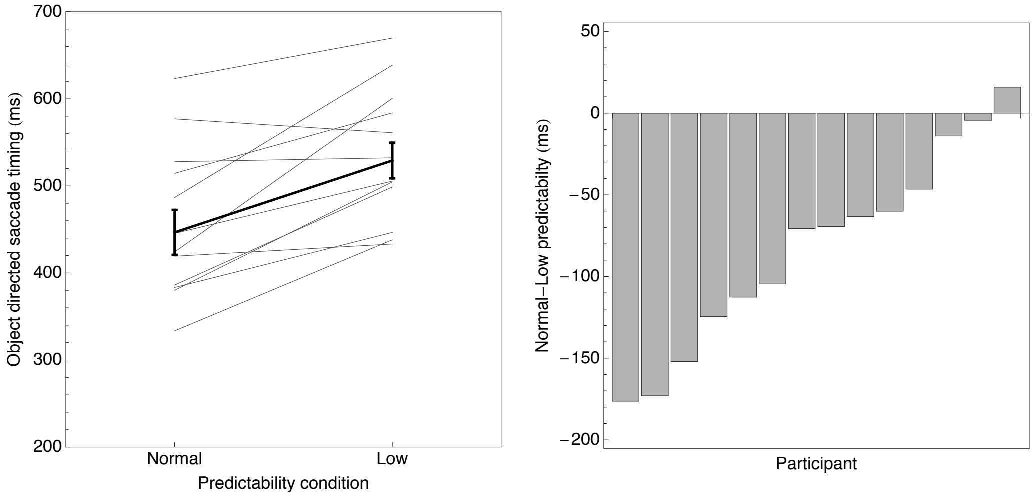

Figure 13: Left panel, object directed saccade timing in the normal and low predictability condition. The gray lines represent the performance of each participant, whereas the black line represents the mean values. The error bars represent the standard error of the mean. Right panel, difference in the timing of object directed saccades between the normal and the low predictability conditions for each participant.

The present results clearly show that when both participants were adapting their own hand movements to the trajectory of the other participant, they could start monitoring their partner behavior later in time than when only one participant had to take care of the whole adjustment by him/herself.

In the normal predictability condition, during the execution of each trial both partners were performing object directed saccades with a very high regularity. To explore the relationship between gazeing behaviors of the two partners we calculated the Pearson's product moment between the timings of the object directed saccades. The correlation coefficients ranged between 0.27 and 0.48. These correlations thus suggest that the partners were adjusting the timings of the object directed saccades with a certain degree of coordination, either both preceding of both delaying the moment in which they were performing the saccade toward the partner's object.

5.4 Discussion

Human gaze behavior has been studied in various natural activities. The allocation of gaze has been studied in tasks such as copying arrangements of blocks (Ballard et al, 1995), making tea (Land et al, 1999), making sandwiches (Hayhoe et al, 2003), driving (Land and Lee, 1994), and other goal-directed behavior. A common finding was that the use of gaze is highly dependent on the task requirements. However, the gaze behavior is principally used to gather visual information to guide movements.

In the present study, we extended the experimental paradigm introduced by Johansson et al (2001) to a human-human interaction setting. Whereas in Johansson et al (2001) study the gaze behavior of a participant that had to move an object and make contact with a target was measured, we measured simultaneously the gaze behavior of two participants that had to move each his/her own object and they had to then make contact with the object of the other participant. This variation introduced the necessity for human-human coordination and cooperation. The contact location was always determined online during the execution of the hand movements and was a consequence of the adaptive behaviors of both participants. To study this aspect further we measured a second condition in which one of the participants was instructed to disregard the hand trajectory of the other participant and just position his/her own object at a location of choice. The other participant had thus to take the lead of the whole adaptive process.

First, we looked where do humans normally look when they interact in a common task together with another human. Participants directed gaze almost exclusively to objects involved in the task. These included their own object, parts of the structure that served as a reference for the hand movement trajectory and the partner's object. The partner's object prior to making contact with it was an obligatory gaze target. Participants performed regularly these object directed saccades. This means that in the final phase of each hand movement the participants monitored each others hand movements to adapt their own movements and successfully reach the common goal.

In the low predictability condition, the gaze behavior of the participant that had to just position the object by disregarding the partner's behavior was quite different. Not having the need to monitor the partner's behavior resulted in an absence of eye movements toward the partner's object. On the other hand, the active participant exhibited the same gaze behavior pattern as in the normal predictability condition. The only difference was in the temporal characteristics of the sequence of gaze movements.

Second, we looked how does the gazing behavior depend on the cooperativeness of the other human. In particular, the timing of the object directed saccades differed significantly between the normal and the low predictability conditions. The saccades were initiated earlier in the low predictability condition. This was an indication that the focus of attention had to be deployed towards the partner's object earlier to extract its trajectory and better predict the contact location. On the contrary, in the normal predictability condition both partners were simultaneously adjusting their trajectories toward a commonly agreed contact location and therefore they could direct their attention to the partner's object at a later stage. This is a clear example of how the brain uses gaze fixations to obtain spatial information for controlling manipulatory actions at the moment in time at which the spatial information is most needed.

The stereotypical gaze behavior seems necessary to establish a closed loop between the two participants that allows a coordinated fine-tuning of the joint interaction. When both participants jointly adapt their behavior for the achievement of the common goal, fewer resources are needed for a successful interaction. The expectations that a human actor has about the cooperation partner influence the deployment of gaze movements and, consequently, the deployment of attentional resources.

References

Anstis SM, Mayhew JW, Morley T (1969) The perception of where a face or television ”portrait” is looking. Am J Psychol 82(4):474–489

Ballard D, Hayhoe M, Pelz J (1995) Memory Representations in Natural Tasks. J Cogn Neurosci 7(1):66–80

Ballard DH, Hayhoe MM, Li F, Whitehead SD (1992) Hand-eye coordination during sequential tasks. Philos Trans R Soc Lond B Biol Sci 337(1281):331–338

Bock SW, Dicke P, Thier P (2008) How precise is gaze following in humans? Vision Res 48(7):946–957

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10(4):433–436

Cline MG (1967) The perception of where a person is looking. Am J Psychol 80(1):41–50

Cornelissen FW, Peters EM, Palmer J (2002) The eyelink toolbox: eye tracking with matlab and the psychophysics toolbox. Behav Res Methods 34(4):613–617

Crostella F, Carducci F, Aglioti SM (2009) Reflexive social attention is mapped according to effector-specific reference systems. Exp Brain Res 197(2):143–151

Driver J, Davis G, Ricciardelli P, Kidd P, Maxwell E, Baron-Cohen S (1999) Gaze perception triggers reflexive visuospatial orienting. Vis Cogn 6(5):509–540

Friesen C, Kingstone A (1998) The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. Psychon Bull Rev 5:490–495

Gibson J, Pick A (1963) Perception of another person’s looking behavior. Am J Psychol 76:386–394

Hayhoe MM, Shrivastava A, Mruczek R, Pelz JB (2003) Visual memory and motor planning in a natural task. J Vis 3(1):49–63

Johansson RS, Westling G, Bäckström A, Flanagan JR (2001) Eye-hand coordination in object manipulation. J Neurosci 21(17):6917–6932

Kuhn G, Benson V (2007) The influence of eye-gaze and arrow pointing distractor cues on voluntary eye movements. Percept Psychophys 69(6):966–971

Kuhn G, Kingstone A (2009) Look away! Eyes and arrows engage oculomotor responses automatically. Atten Percept Psychophys 71(2):314–327

Land M, Mennie N, Rusted J (1999) The roles of vision and eye movements in the control of activities of daily living. Perception 28(11):1311–1328

Land MF, Lee DN (1994) Where we look when we steer. Nature 369(6483):742–744

Lobmaier JS, Fischer MH, Schwaninger A (2006) Objects capture perceived gaze direction. Exp Psychol 53(2):117–122

Mansfield E, Farroni T, Johnson M (2003) Does gaze perception facilitate overt orienting?. Vis Cogn 10(1):7–14

Pelli DG (1997) The videotoolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10(4):437–442

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32(1):3–25

Ricciardelli P, Bricolo E, Aglioti SM, Chelazzi L (2002) My eyes want to look where your eyes are looking: exploring the tendency to imitate another individual's gaze. Neuroreport 13(17):2259–2264

Ricciardelli P, Betta E, Pruner S, Turatto M (2009) Is there a direct link between gaze perception and joint attention behaviours? Effects of gaze contrast polarity on oculomotor behaviour. Exp Brain Res 194(3):347–357

Schwaninger A, Lobmaier JS, Fischer MH (2005) The inversion effect on gaze perception reflects processing of component information. Exp Brain Res 167(1):49–55

Symons LA, Lee K, Cedrone CC, Nishimura M (2004) What are you looking at? Acuity for triadic eye gaze. J Gen Psychol 131(4):451–469